CROATICA CHEMICA ACTA CCACAA **75** (4) 935–944 (2002)

ISSN-0011-1643
CCA-2843

Original Scientific Paper

Machine Learning Based Analysis of Biochemical and Morphologic Parameters in Patients with Dialysis Related Amyloidosis

Igor Barišić,[a] *Vladimir Wilhelm,*[b] *Nikola Štambuk,*[c,*] *Ksenija Karaman,*[d] *Stipan Janković,*[a] *Paško Konjevoda,*[c] *and Biserka Pokrić*[c]

[a]*Department of Diagnostical and Interventional Radiology, Clinical Hospital Split, Šoltanska 1, 21000 Split, Croatia*

[b]*Division of Nephrology, Department of Internal Medicine, Clinical Hospital Split, Šoltanska 1, 21000 Split, Croatia*

[c]*Rudjer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*

[d]*Department of Ophthalmology, Clinical Hospital Split, Šoltanska 1, 21000 Split, Croatia*

Received July 24, 2002; accepted September 6, 2002

Dialysis related amyloidosis is the accumulation and deposition of β_2-microglobulin derived fibrils in bones and joints, due to insufficient elimination during therapy or slowly progressing renal failure. The aim of this study was to analyse biochemical, morphologic and anamnestic parameters that may be relevant for the onset and developement of dialysis related amyloidosis. In addition to standard statistical procedures, we also applied the machine-learning based methods of data mining to quantify the risk factors for asymptomatic patients. Extraction of risk factors for the onset of the dialysis related amyloidosis syndrome could enable the clinician to predict the symptoms and consider medical procedures to prevent the onset of the disease. The C4.5 machine learning algorithm extracted a simple and highly accurate tree for discrimination of asymptomatic and symptomatic patients suffering from dialysis related amyloidosis. It remains an open question if our

* Author to whom correspondence should be addressed. (E-mail: stambuk@rudjer.irb.hr)

findings may contribute to the problem of accurately predicting the onset of dialysis related arthropathy in the asymptomatic patient group.

Key words: dialysis, amyloidosis, biochemistry, morphologic parameters, shoulder, knee, symptoms.

INTRODUCTION

Dialysis related amyloidosis is defined as the accumulation and deposition of β_2-microglobulin derived fibrils, especially in bones and joints, due to insufficient elimination during therapy. The syndrome has also been reported in patients with slowly progressing renal failure who had never been dialysed.[1–3]

Amyloidosis in dialysis patients was first linked to osteoarticular syndrom 17 years ago.[1] The clinical manifestations were dialysis arthropaty and carpal tunnel syndrom. Chronic arthropaty is the term that has also been reported by many authors.[4] Clinical signs of the dialysis related amyloidosis are not specific in some cases. Typical radiological signs such as amyloid bone cysts are often late events,[4] and capsulosynovial swelling precedes the development of characteristic bone cysts, so ultrasound is the method of choice for observing the early signs.[5–13]

Shoulder is the joint most frequently involved and changes are present mainly in tendons and bursae. Opposite results are obtained in the knee where changes are present primarily in the articular structures.[4]

Many factors are thought to be related to the syndrome, *e.g.,* age, duration of dialysis, remaining renal function, and related inflammatory disorders.[2,3,14–17]

The aim of this study was to extract radiologic, laboratory and anamnestic parameters that may be relevant for the onset and developement of dialysis related amyloidosis. In addition to standard statistical procedures, we also applied the machine-learning based methods of data mining to quantify the risk factors for asymptomatic patients.

Extraction of risk factors for the onset of the dialysis related amyloidosis syndrome could enable clinicians to predict the symptoms and consider medical procedures to prevent the onset of the disease.

MATERIAL AND METHODS

Patients and Imaging Examinations

Real time sonography of the shoulders and knees was performed in 40 patients receiving chronic haemodialysis as treatment of terminal renal failure. Linear trans-

ducers of 7.5 and 10 Mhz were used. Plain radiographs of the shoulders and knees in standard positions were taken simultaneously.

The group of patients consisted of 26 males and 14 females. Their mean age was 55.8 years (range 20–74 years). The mean duration of haemodialysis was 79.3 months (range 12–211 months).

Criteria for clinical diagnosis of dialysis related amyloidosis were: persistent pain and stiffness in both shoulders and knees lasting for more than 6 weeks and restriction of movements in various degrees.[5,6,18]

Tendons were examined in 80 shoulders, evaluating the thickening of the rotator cuff, of especially m. supraspinatus tendon. The thickness of supraspinatus was measured in both longitudinal and transversal views, in neutral position of the shoulder and in the adduction, hyperextension and internal rotation of the arm. The mean values between transversal and longitudinal views in both positions were calculated as follows:

(i) Transverse view: superior of the head of humerus, just above the bicipital groove, taken mid-point between symmetrical lateral narrowing of the tendon;

(ii) Longitudinal view: measured at the point where tendon emerges beneath acromial shadow.[6]

During examination, texture of the tendons, especially inhomogencity, hyperechoic amyloid deposits and small calcium deposits were observed.[6,19]

The knees were evaluated considering the presence of joint effusion and Baker's cysts. Shoulder measurements of 40 healthy volunteers and 60 patients taken during other types of examinations in the USA were taken as referent points. Conventional plain radiographs were analysed looking for the presence of calcifications of the tendons and bursae.

Laboratory Examinations

Thirty-eight laboratory tests of peripheral blood biochemical parameters presented in Table I were done periodically as a routine procedure during a one year prospective trial.

Data Analysis

Data were analysed by means of two software packages. STATISTICA for Windows version 5.0 (http://www.StatSoft.com) was used for the estimation of Hotelling's T test and chi-square tests.[20] Machine-learning analysis of biochemical and imaging examinations was done with respect to the presence of characteristic joint symptoms of dialysis related amyloidosis. The analysis was done by means of the C4.5 (J48) machine learning classifier, with Weka (Waikato Environment for Knowledge Analysis) software version 3.1.7.[20–22] Weka is freely available at the World Wide Web address http://www.cs.waikato.ac.nz/ml/weka.

C4.5 Machine-learning Program

The C4.5 program is a successor of the basic ID3 decision tree learning algorithm.[22–24] C4.5 defines the possible decision tree by means of a hill-climbing search based on the statistical property measure called information gain.[22–24] The elements

of the tree generated by C4.5 are either leaves or decision nodes.[22–24] The leaf shows a class and the decision node specifies the test to be implemented on an attribute value, with one branch and a subtree for each possible result of the test.[22–24] The starting node is the root node and the tree is used to predict a case by starting at the root and moving through the tree until the leaf is encountered.[22–24] For any tree, all paths lead to a leaf corresponding to a decision rule that is a logical conjunction of various tests.[22–24]

RESULTS AND DISCUSSION

We compared the results of biochemical tests and morphologic parameters in patients with symptoms associated with dialysis related amyloidosis and asymptomatic patients (Table I). In addition to elevated serum β_2-microglobulin levels, previously reported by several authors, we also detected a statistically significant rise in CRP, typical of chronic inflammation. Drop of albumin values in patients with dialysis related amyloidosis in Table I may be also linked to chronic inflammatory events. It is worth mentioning that in symptomatic patients the duration of dialysis treatment was significantly longer and that morphologic parameters in the knees showed no significant differences between the groups.

As regards, morphologic parameters, significant differences were recorded for sonographic inhomogeneity of the rotator cuff and RTG calcifications of the rotator cuff (Table II). Thickening of the rotator cuff (Figure 1), especially the supraspinatus one was reported to be present in patients receiving a long-term dialysis.[4,6,10,12] However, in this study, the sonographic inhomogeneity of the rotator cuff, instead of the thickening of the rotator cuff, was extracted as a parameter relevant for the diagnosis (Figure 2, Table II). Inhomogeneity parameter that we found to be relevant for the discrimination of symptomatic and asymptomatic patients (Figure 2, Table II) is often related to different pathological findings, including tissue deposits or microtrauma with bleeding, degeneration, fibrosis or inflammation.[4] Consequently, the inhomogeneity of the rotator cuff seems to be typical of the affected shoulder findings in dialysis related amyloidosis (Table II). RTG calcifications of the rotator cuff are not necessarily associated with amyloidosis because they appear in other dialysis complications and in other pathologic syndromes such as impingement syndrome or calcific tendinitis.

In order to extract the algorithm linking the biochemical and morphologic parameters to the diagnosis of the dialysis related amyloidosis, we have applied Quinlan's C4.5 machine learning algorithm.[22–24] The decision tree based on the C4.5 algorithm is presented in Figure 2. Machine learning algorithm extracted the sonographic inhomogeneity (SI) of the right shoulder rotator cuff, duration of dialysis and β_2-microglobulin serum levels as the parameters that with 97.4% accuracy discriminate symptomatic from asymptomatic

TABLE I

The results of biochemical tests in patients with dialysis related amyloidosis (positive) and asymptomatic dialysis patients (negative)

Values in peripheral blood	Mean negative	Mean positive	SD negative	SD positive	*t*-value	*p*
RBC count $\times 10^{12}$ / dm^{-3}	3.41	3.46	0.51	0.71	–0.228	0.8210
Hemoglobin, Hb / g dm^{-3}	96.75	104.40	16.90	17.52	–1.356	0.1833
Hematocrit, Hct	1.28	0.32	4.84	0.06	0.764	0.4499
MCV	86.59	91.47	5.52	5.31	–2.730	0.0096
Reticulocyte count / %	2.23	2.45	0.71	0.95	–0.799	0.4292
Platelet count $\times 10^{9}$ / dm^{-3}	195.00	184.40	77.04	60.03	0.453	0.6532
WBC count $\times 10^{9}$ / dm^{-3}	6.41	6.06	1.72	2.03	0.558	0.5804
Urea nitrogen blood, BUN / mg dm^{-3}	27.72	28.51	7.34	5.70	–0.358	0.7226
Creatinine / μmol dm^{-3}	881.42	840.67	195.05	139.67	0.703	0.4866
Uric acid / μmol dm^{-3}	418.54	448.80	67.28	64.32	–1.389	0.1731
Bilirubin / μmol dm^{-3}	11.48	12.76	3.20	7.47	–0.740	0.4640
Glucose / mmol dm^{-3}	6.33	6.36	2.37	2.02	–0.036	0.9714
Cholesterol, total / mmol dm^{-3}	4.53	4.42	0.82	1.24	0.345	0.7322
Proteins, total / g dm^{-3}	73.04	73.53	4.97	4.19	–0.318	0.7520
Albumin / g dm^{-3}	43.13	40.80	3.60	2.91	2.104	0.0422
Alkaline phosphatase / U dm^{-3}	81.75	79.53	51.76	39.87	0.141	0.8883
Aspartate amino-transferase, AST / U dm^{-3}	13.29	12.20	9.52	5.97	0.397	0.6936
Alanine amino-transferase, ALT / U dm^{-3}	16.50	13.00	16.81	10.36	0.723	0.4741

TABLE I (cont.)

γ-Glutamyltransferase, GGT / U dm^{-3}	36.25	39.73	36.97	42.92	–0.269	0.7894
Lactate dehydrogenase, LD / U dm^{-3}	274.75	282.27	43.44	80.88	–0.378	0.7075
Calcium / mmol dm^{-3}	2.52	2.43	0.40	0.24	0.715	0.4792
Phosphorus / mmol dm^{-3}	1.67	1.45	0.45	0.50	1.428	0.1616
Magnesium / mmol dm^{-3}	1.12	1.04	0.14	0.18	1.414	0.1660
Iron / μmol dm^{-3}	12.25	10.57	6.34	3.76	0.927	0.3597
Iron-binding capacity, UIBC / μmol dm^{-3}	27.32	24.53	8.81	4.15	1.144	0.2598
Sodium / mmol dm^{-3}	137.75	137.80	2.23	4.57	–0.046	0.9637
Potassium / mmol dm^{-3}	5.65	5.94	0.84	0.96	–1.009	0.3196
Chloride, Cl / mmol dm^{-3}	99.04	93.43	2.99	23.50	1.164	0.2520
Ferritin / μg dm^{-3}	626.83	910.67	822.74	558.97	–1.175	0.2477
IgG / g dm^{-3}	11.40	12.76	2.89	3.58	–1.279	0.2089
IgA / g dm^{-3}	1.95	2.36	1.04	1.62	0.965	0.3409
IgM / g dm^{-3}	1.16	1.00	0.66	0.58	0.756	0.4542
C-reactive protein, CRP / mg dm^{-3}	13.83	25.27	8.15	19.01	–2.603	0.0132
β_2-Microglobulin / mg dm^{-3}	16.96	24.28	7.70	8.18	–2.822	0.0008
Urea nitrogen blood post. × 10 / mg dm^{-3}	11.87	10.33	3.42	2.76	1.468	0.1506
Creatinine after dialysis / μmol dm^{-3}	411.25	394.20	104.83	76.97	0.544	0.5898
Sodium after dialysis / mmol dm^{-3}	141.67	141.33	2.32	3.60	0.353	0.7261
Potassium after dialysis / mmol dm^{-3}	4.48	4.55	0.69	0.65	–0.304	0.7627

valid *n* negative = 24
valid *n* positive = 15

TABLE II

Morphologic parameters in patients with dialysis related amyloidosis and in asymptomatic patients

Patients	Sonographic inhomogeneity of the rotator cuff / %[a]	RTG calcifications of the rotator cuff / %[b]
Symptomatic	96.7	100
Asymptomatic	27.3	63.3

[a] $p = 0.00001$ (Yates corrected $\chi^2 = 24.71$).
[b] $p = 0.00430$ (Yates corrected $\chi^2 = 8.15$).

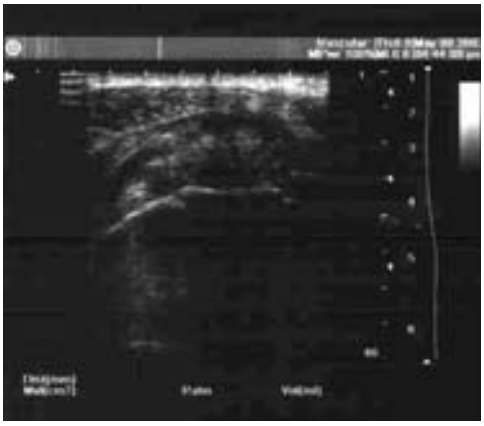

Figure 1. Thickening and inhomogeneity of the rotator cuff.

patients (Figure 2). Leave-one-out crossvalidation testing of the procedure's accurracy confirmed its validity by a very high overall classification of 87.18% (Figure 2). The results of Table II are in line with the previously discussed results of Table I and Figure 2. In addition to the biochemical parameter of serum β_2-microglobulin and the morphologic parameter of rotator cuff inhomogeneity, thought to be involved in the pathophysiological events of the dialysis related amyloidosis, the tree also extracted the duration of the dialysis as the essential parameter that discriminates the groups of symptomatic from asymptomatic patients (Figure 2). The latter is consistant with literature findings.[1,4,5–7,12]

The C4.5 machine learning algorithm enabled us to extract a simple and highly accurate tree for the discrimination of asymptomatic and symptomatic patients suffering from dialysis related amyloidosis. It remains an open question if our findings may contribute to the problem of accurately predicting the onset of dialysis related arthropathy in asymptomatic patients' group.

a)

SI of right rotator cuff ≤ 0: negative (13.5)
SI of right rotator cuff > 0
| DURATION ≤ 18: negative (6.54)
| DURATION > 18
| | β_2-M ≤ 15.9: negative (4.96/1.0)
| | β_2-M > 15.9: positive (14.0)

Number of leaves: 4
Size of the tree: 7
Time taken to build the model: 0.05 seconds

b)

=== Evaluation on the training set (Summary) ===

Correctly classified instances	38	97.4359%
Incorrectly classified instances	1	2.5641%
Kappa statistic	0.9451	
Mean absolute error	0.0409	
Root mean squared error	0.1406	
Relative absolute error	8.6262%	
Root relative squared error	28.8963%	
Total number of instances	39	

=== Detailed accuracy by class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0.067	0.96	1	0.98	negative
0.933	0	1	0.933	0.966	positive

=== Confusion matrix ===

a	*b*	<-- classified as
24	0	*a* = negative
1	14	*b* = positive

=== Confusion matrix === LEAVE-ONE-OUT CROSSVALIDATION

a	*b*	<-- classified as
22	2	*a* = negative
3	12	*b* = positive

Figure 2. (a) C4.5 decision tree analysis of biochemical and morphologic parameters relevant for the diagnosis of dialysis related amyloidosis; (b) Statistical evaluations and procedure cross validation.

REFERENCES

1. T. Bardin, D. Kuntz, J. Zingraff, M. Voisin, A. Zelmar, and J. Lansman, *Arthritis Rheum.* **28** (1985) 1052–1058.
2. J. J. Zingraff, L. H. Noel, and T. Bardin, *N. Engl. J. Med.* **323** (1990) 1070–1071.
3. M. Jadoul, C. Garbar, and H. Noel, *Kidney Int.* **51**(1997) 1928–1932.
4. G. Coari, A. Iagnocco, S. Maggi, M. Bracci, A. De Cata, M. Mastantuomo, M. Larciprete, and S. Persichetti, *Eur. Radiol.* **6** (1996) 890–894.
5. D. Sethi, T. C. Morgan, E. A. Brown, N. R. Cary, C. C. Erhardt, M. Pazanas, R. N. Maini, D. F. Woodrow, and P. E. Gower, *Q. J. Med.* **77** (1990) 1061–1082.
6. L. P. Mc Mahon, J. Radford, and J. K. Daeborn, *Clin. Nephrol.* **35** (1991) 227–232.
7. M. H. Kurer, R. A. Baillod, and J. C. Madgwick, *J. Bone Joint Surg.* **73** (1991) 271–276.
8. L. Badrati, B. Balbi, A. Rocchi, R. Bonsanto, D. Docci, C. Capponcini, C. Feletti, and M. Mughetti, *Radiol. Med.* **81** (1991) 234–237.
9. R. Nessi, S. Bolzoni, D. Brancaccio, and C. Uslenghi, *Radiol. Med.* **85** (1993) 252–256.
10. M. Jadoul, J. Malghem, B. van de Berg, and C. van Ypersele de Strihou, *Kidney Int. Suppl.* **41** (1993) 106–110.
11. S. Negi, Y. Kita, K. Uchita, and T. Abe, *Nippon Jinzo Gakkai Shi* **37** (1995) 29–34.
12. E. Cardinal, K. A. Buckwalter, E. M. Braunstein, D. Raymond-Tremblay, and M. D. Bensos, *Am. J. Roentgenol.* **166** (1996) 153–156.
13. U. Rapp-Bernhardt, H. Milbradt, T. M. Bernhardt, and W. Dohring, *Ultraschall Med.* **18** (1997) 91–94.
14. Y. S. C. Van, M. Jadoul, J. Malghem, B. Maldaague, and J. Jamart, *Kidney Int.* **39** (1991) 1012–1019.
15. T. Miyata, R. Inagi, Y. Lida, M. Sato, N. Yamada, and O. Oda, *J. Clin. Invest.* **93** (1994) 521–528.
16. T. Miyata, Y. Lida, Y. Ueda, T. Shizato, H. Seo, V. M. Monnier, and K. Maeda, *Kidney Int.* **49** (1996) 538–550.
17. I. L. Noronha, C. K. Fujihara, and R. Zatz, *Nephrol. Dial. Transplant.* **17** (2002) 363–368.
18. N. P. Hurst, R. van der Berg, A. Disney, M. Alcoc, L. Albertin, and M. Green, *Ann. Rheum. Dis.* **48** (1989) 409–420.
19. J. Kay, C. B. Bensosn, and S. Lester, *Arthritis Rheum.* **35** (1992) 926–932.
20. N. Štambuk and P. Konjevoda, *Int. J. Quant. Chem.* **84** (2001) 13–22.
21. I. H. Witten and E. Frank, *Data Mining*, San Francisco, Morgan Kaufmann Publishers, 2000, pp.169–170.
22. S. Seiwerth, N. Štambuk, P. Konjevoda, N. Mašić, A. Vasilj, M. Bura, I. Klapan, S. Manojlović, and D. Đanić, *J. Chem Inf. Comput. Sci.* **40** (2000) 545–549.
23. J. R. Quinlan, *Machine Learning* **1** (1986) 81–106.
24. T. M. Mitchell, *Machine Learning*, Singapore, McGrow-Hill, 1997, pp. 52–80.

SAŽETAK

Analiza biokemijskih i morfoloških pokazatelja u pacijenata s amiloidozom u dijalizi zasnovana na strojnom učenju

Igor Barišić, Vladimir Wilhelm, Nikola Štambuk, Ksenija Karaman, Stipan Janković, Paško Konjevoda i Biserka Pokrić

Amiloidoza povezana s dijalizom jest nakupljanje i odlaganje fibrila β_2-mikroglobulina u kostima i zglobovima zbog nedovoljne eliminacije tijekom terapije ili sporo progredirajućeg zatajenja bubrega. Cilj rada bio je utvrditi biokemijske, morfološke i anamnestičke pokazatelje koji bi mogli biti značajni za početak i razvoj amiloidoze koja nastaje tijekom dijalize. Uz standardne statističke metode, u radu su za analizu podataka uporabljene i metode strojnog učenja, radi kvantifikacije rizičnih faktora za asimptomatske pacijente. Izdvajanje rizičnih faktora za pojavu sindroma amiloidoze u dijalizi moglo bi kliničaru omogućiti predviđanje početka simptoma i primjenu preventivnih medicinskih postupaka u cilju sprječavanja početka bolesti. S pomoću algoritma C4.5 izdvojeno je jednostavno stablo visoke točnosti za razlučivanje asimptomatskih od simptomatskih bolesnika s amiloidozom u dijalizi. Ostaje otvoreno pitanje mogu li naši rezultati pridonijeti točnijem predviđanju početka artropatije povezane s dijalizom u skupini asimptomatskih bolesnika.